

What is claimed is:

1. A method for testing output quality from a data extraction process, comprising:

receiving input data containing information to be inserted into a database;

dividing the input data into a plurality of batches such that a subset of the input data is
5 duplicated among the plurality of batches;

distributing the plurality of batches to a plurality of data entry clerks, wherein each data entry clerk processes one of the plurality of batches and converts data from the batch into worked data;

receiving the worked data from each of the plurality of data entry clerks; and

inspecting the subset of the worked data corresponding to the duplicated subset of the input data to determine the accuracy of the subset of worked data.

2. The method for testing output quality from claim 1, wherein the step of inspecting predicts the quality of the worked data.

3. The method for testing output quality from claim 1, wherein the subset of the input data duplicated among the batches is based on a sampling plan.

4. The method for testing output quality from claim 1, further comprising repeating the steps of dividing, distributing, receiving and inspecting, if a desired level of accuracy is not reached.

5. The method for testing output quality from claim 1, further comprising adjusting
20 the desired level of accuracy based on inspecting the subset of the worked data.

6. The method for testing output quality from claim 1, wherein the step of inspecting the subset of the worked data comprises:

identifying the subset of the worked data resulting from the duplicated subset of the input data;

comparing entries made by each of the plurality of data clerks on the subset of the worked data; and

flagging the entries that differ.

7. The method for testing output quality from claim 1, wherein the step of inspecting the subset of the worked data comprises: accepting the worked data for submission to a database if the desired level of accuracy is met and rejecting the worked data for submission to the database if the desired level of accuracy is not met.

8. The method for testing output quality from claim 1, wherein the input data is a plurality of technical product data sheets.

9. The method for testing output quality from claim 1, wherein the steps of dividing, distributing, receiving and inspecting are accomplished with a computer system.

10. A data extraction tool implemented on a computer, the tool comprising:

a first receiver unit for receiving input data containing information to be inserted into a database;

a data divider unit for dividing the input data into a plurality of batches such that a subset of the input data is duplicated among the plurality of batches;

a distributor unit for distributing the plurality of batches to a plurality of data entry clerks, wherein each data entry clerk processes one of the plurality of batches and converts data from the batch into worked data;

a second receiver unit for receiving the worked data from each of the plurality of data entry clerks; and

an inspector unit for inspecting the subset of the worked data corresponding to the duplicated subset of the input data to determine the accuracy of the subset of worked data.

11. The data extraction tool implemented on a computer from claim 10, wherein the inspector unit predicts the quality of the worked data.

12. The data extraction tool implemented on a computer from claim 10, wherein the subset of the input data duplicated among the batches is based on a sampling plan.

5 13. The data extraction tool implemented on a computer from claim 10, further comprising reworking the batch using the distributor unit, second receiver unit, and inspector unit, if a desired level of accuracy is not reached.

14. The data extraction tool implemented on a computer from claim 10, further comprising adjusting the desired level of accuracy based on the inspector unit inspecting the subset of the worked data.

15. The data extraction tool implemented on a computer from claim 10, wherein the inspecting of the subset of the worked data performed by the inspector unit comprises:

identifying the subset of the worked data resulting from the duplicated subset of the input data;

comparing entries made by each of the plurality of data clerks on the subset of the worked data; and

flagging the entries that differ.

16. The data extraction tool implemented on a computer from claim 10, wherein the inspecting of the subset of the worked data performed by the inspector unit comprises:
20 accepting the worked data for submission to a database if the desired level of accuracy is met and rejecting the worked data for submission to the database if the desired level of accuracy is not met.

17. The data extraction tool implemented on a computer from claim 10, wherein the input data is a plurality of technical product data sheets.

18. A computer program for a data extraction tool, the computer program embodied on a computer readable medium for execution by a computer, the computer program comprising:

a code segment that receives input data containing information to be inserted into a database;

a code segment that divides the input data into a plurality of batches such that a subset of the input data is duplicated among the plurality of batches;

a code segment that distributes the plurality of batches to a plurality of data entry clerks, wherein each data entry clerk processes one of the plurality of batches and converts data from the batch into worked data;

a code segment that receives the worked data from each of the plurality of data entry clerks; and

a code segment that inspects the subset of the worked data corresponding to the duplicated subset of the input data to determine the accuracy of the subset of worked data.

19. The computer program for a data extraction tool from claim 18, wherein the code segment that inspects the data predicts the quality of the worked data.

20. The computer program for a data extraction tool from claim 18, wherein the subset of the input data duplicated among the batches is based on a sampling plan.

21. The computer program for a data extraction tool from claim 18, further comprising reworking the batch using the code segment that distributes, the code segment that receives, and the code segment that inspects, if a desired level of accuracy is not reached.

22. The computer program for a data extraction tool from claim 18, further comprising adjusting the desired level of accuracy based the code segment that inspects inspecting the subset of the worked data.

23. The computer program for a data extraction tool from claim 18, wherein the step of inspecting performed by the code segment that inspects comprises:

identifying the subset of the worked data resulting from the duplicated subset of the input data;

5 comparing entries made by each of the plurality of data clerks on the subset of the worked data; and

flagging the entries that differ.

24. The computer program for a data extraction tool from claim 18, wherein the step of inspecting performed by the code segment that inspects comprises: accepting the worked data for submission to a database if the desired level of accuracy is met and rejecting the worked data for submission to the database if the desired level of accuracy is not met.

25. The computer program for a data extraction tool from claim 18, wherein the input data is a plurality of technical product data sheets.